Association for Information Systems
AIS Electronic Library (AISeL)

ICIS 2002 Proceedings

International Conference on Information Systems (ICIS)

December 2002

The Virtual Commons: Why Free-Riding Can Be Tolerated in File Sharing Networks

Ramayya Krishnan
Carnegie Mellon University

Michael Smith
Carnegie Mellon University

Zhulei Tang
Carnegie Mellon University

Rahul Telang
Carnegie Mellon University

THE VIRTUAL COMMONS: WHY FREE-RIDING CAN BE TOLERATED IN FILE SHARING NETWORKS

Ramayya Krishnan
H. John Heinz III School of Public Policy and Management
Carnegie Mellon University
Pittsburgh, PA USA
rk2x@andrew.cmu.edu

Michael D. Smith
H. John Heinz III School of Public Policy and Management
Carnegie Mellon University
Pittsburgh, PA USA
mds@andrew.cmu.edu

Zhulei Tang
Graduate School of Industrial Administration
Carnegie Mellon University
Pittsburgh, PA USA
zhuleit@andrew.cmu.edu

Rahul Telang
H. John Heinz III School of Public Policy and Management
Carnegie Mellon University
Pittsburgh, PA USA
rtelang@andrew.cmu.edu

Abstract

Peer-to-peer networks have emerged as a popular alternative to traditional client-server architectures for the distribution of information goods. Recent academic studies have observed high levels of free-riding in various peer-to-peer networks, leading some to suggest the imminent collapse of these communities as a viable information sharing mechanism. Our research develops an analytic model to analyze the behavior of P2P networks in the presence of free-riding. In contrast to previous predictions we find that P2P networks can operate effectively in the presence of significant free-riding. In future work we plan to explore how much peer-to-peer network performance could be improved if free-riding were eliminated and discuss both the costs and benefits of managerial mechanisms to limit free-riding.

1 INTRODUCTION

Peer-to-peer (P2P) networks have emerged recently as an alternative to traditional client-server networks for the distribution of information goods. Popular P2P sites, including Napster, Kazaa, and Morpheus Music City, have attracted millions of global users to share songs, films, software, and computer games. Content shared in these networks is typically unlicensed copyrighted files. However, the P2P network architecture can be used to share non-copyright infringing content. For example, the subscription-based Napster service launched in early 2002 used DRM technology within its P2P architecture to protect copyright holders (Krishnan et al. 2002). Likewise, the P2P architecture is gaining popularity for applications including distributed computing (e.g., SETI@Home), enterprise knowledge sharing (e.g., Bad Blue), and user collaboration (e.g., Groove Networks).

While P2P networks vary in their architectural design, in all P2P networks, files are transferred directly between the computers of users (a.k.a. peers) connected to the network. Further, once these files have been delivered, the user accessing the file, by default, becomes a provider of that content. Thus, in an ideal case, the provision of content on the network will scale to match the level of demand for the content. This characteristic also means that P2P networks can be modeled in the context of the economic concept of public goods and club goods. In contrast to private goods, public goods have the characteristics of non-excludability in supply (individuals can't be excluded from consuming the product) and non-rivalry in demand (one individual's consumption does not diminish another user's value of the product) (Hardin 1968). Clean air is a typical example of a public good. Club goods are goods that are excludable in supply but non-rival in demand (Buchanan 1965).

In the ideal case, P2P networks will exhibit both non-excludability (information is made available to all members of the network) and non-rivalry (consumption by one user doesn't decrease network download possibilities in the absence of free-riding). However, in the presence of free-riding, P2P networks will exhibit levels of rivalry (Asvanund et al. 2001), which distinguishes them from either public goods or club goods.

Applying the public and club goods literatures to P2P networks, we seek to explain two interesting observations regarding P2P network performance. First, researchers have observed high levels of free-riding in P2P networks. While users share downloaded content, by default, they can and frequently do turn off this feature to economize on their own private allocation of bandwidth. For example, Adar and Huberman (2000) find that 70 percent of the users in the Gnutella network do not contribute content to the system and the top 1 percent of users contribute over 50 percent of the total amount of content. Asvanund et al. find evidence of free-riding in the context of OpenNap networks. Traditional economic theory predicts that the free-rider problem causes inefficient private provision of public goods, and calls for a central intervention to remedy this problem. In the context of P2P networks, Adar and Huberman observe, "Free-riding leads to degradation of the system performance…if this trend continues copyright issues might become moot compared to the possible collapse of such systems."

In contrast, however, the second observed characteristic of P2P networks is that they appear to persist in spite of these high levels of free-riding. For example, according to a CNN report, the number of P2P sites totals nearly 38,000, up 535 percent in the past year.

Given these two empirical facts, an important question is: How can these networks survive in spite of widespread free-riding? Some researchers have proposed membership rules to encourage peers to share their content (Golle et al. 2001). However, identity can be easily changed on the Internet and this would make it difficult to enforce membership rules. Further, the unique properties of information goods and the feature of P2P protocols preclude the direct application of results from the public goods literature. A careful investigation of peer behavior is needed to develop policy or technical recommendations to understand free-riding in P2P networks. This is the focus of the paper.

2 LITERATURE REVIEW AND CONTRIBUTIONS

Previous research on P2P networks suggests that they have the properties of public and club goods (Asvanund et al. 2002): non-rivalry in demand and/or non-excludability in supply. The major departure of the provision of content in a P2P network from the provision of traditional public goods relies on the fact that consumers also become suppliers by sharing the content they download with other users. This is significant because, in the public goods literature, solutions to free-riding typically involve limiting access to the shared resource. However, in the context of P2P goods, providing full access to the shared resource is an optimal outcome so long as users share their downloaded content with others. In effect, the zero cost of replication of information goods means that consumption of the shared resource can be allowed, even encouraged, so long as users share their downloaded content with other users. Effectively, downloading increases the content available to others.

However, users may not choose to share due to the impact of sharing on their private utility (e.g., their ability to download content from other users). We analyze network behavior in this case by showing that, in many common situations, the users' decision to maximize their private utility will also support the socially optimal outcome. In essence, users realize that by providing content, they can mitigate congestion of the server from which they demand content. This increases the incentive of users to share content in the same way that users have an incentive at a "pot luck" dinner to provide desirable meals.

The main contributions of our study are as follows. First, we develop a game theory framework to model individual users' behavior in P2P networks. Second, we examine peers' incentive to provide content in a context that differs slightly from the traditional provision of public goods. Third, our prediction that P2P networks can persist despite free-riding provides important managerial implications to rights holders and entrepreneurs.

3 CURRENT FINDINGS

3.1 Basic Framework for the Model

In our model, we consider a network where each user has a single endowment of unique content. Each user independently decides whether to share their content based on their private utility. Sharing implies a cost by reducing the user's private bandwidth available for downloading. However, sharing reduces the traffic other peers place on other nodes in the network, thus increasing the user's private utility. The sharing decision serves as a means to redistribute traffic in our model. It provides a rational incentive for some users to share content even without altruism or other external incentives. Each user compares the benefit and cost to formulate their own optimal strategy.

We restrict our attention to the case of perfect information: at each move in the game, the user knows the full history of the play of the game thus far. The perfect information case easily extends to cases of imperfect information.

Consider the following game. In the first period, n users join the network simultaneously. Each user starts with one unit of unique content. Further, each user demands one unit of content randomly from other n-1 users and each user independently decides whether or not to share their own content. In our model, as in most P2P networks, sharing is a binary decision. Users either share all their content files or none of them. If the user shares their own content, whoever desires the content can successfully download it. If the user does not share, content can still be downloaded from other users. Users who download content from others, but do not share their content are referred to as free-riders.

Let $i \in \{SH\}$ be the original set of sharers and $j \in \{NS\}$ be the set of non-sharers (a.k.a. free-riders). Denote g as a representative user. Sharing incurs lump-sum cost c_g. The cost term can be interpreted either in terms of the added congestion sharing will place on the node or in terms of a probability of being prosecuted for copyright violations (in the case that the users were sharing unlicensed copyrighted content). Each user gets value v_g from each content that is downloaded. v_g and c_g are privately known to the user and differ across users. Each user's strategy is $S_g : (v_g, c_g) \rightarrow \{0,1\}$, where 1 stands for sharing, and 0 for non-sharing. An equilibrium of this game is a set of strategies $\{S_g\}$ such that each strategy is optimal given other users' strategies.

Optimality for a user g requires that S_g solve the problem

$$\underset{Sg\in[0,1]}{Max}\left[f\left(\frac{n}{k}\right)\right]v_g - S_g c_g \text{, where } k = \sum_{g=1}^{n} S_g^* . \quad (1)$$

where S^* is the user's optimal strategy.

Notice that $f(\frac{n}{k})$ is the probability that a user can download the content desired, n is the total number of users, and k is the number of sharers. The intuition is if n users divide k units of content, on average each user will get $\frac{k}{n}$ unit of content. Conversely, for each unit of content, the average number of users demanding it is $\frac{n}{k}$. Therefore $\frac{n}{k}$ is expected number of hits on each node. We discuss the derivation in detail in Appendix A.

We assume $f(.)$ is concave in k. That is, the probability to get one piece of content increases with the amount of available content but at a decreasing rate. We first show that in a network of n users there exists equilibrium such that i = 1,……,$k(0 \le k \le n)$ users share their content while the remaining users do not.

Proposition 1: There exists a pure strategy equilibrium such that out of n original users, k ($0 \le k \le n$) users decide to share their content while the rest do not.

Proof: (See Appendix B for proof of proposition.)

Proposition 1 implies that there exist some users who are willing to share content. Thus, the network does not collapse in spite of free-riding. There are two possible reasons. First, in equilibrium, they are better off than non-sharing. Second, when they decide to share their content, they believe that some other users might make the same decision. It is important to note that there are multiple pure strategy equilibria in this game. Thus, it may be difficult to ascertain beforehand which k users might decide to share. But over time, the users will establish a mechanism where some of them may share in a period while the rest free-ride.

3.2 New Users

The first result establishes the existence of free-riding as a stable equilibrium behavior in P2P networks for an initial set of users. What happens as new users join the network? We assume that the new users observe the number of users in {SH} and {NS} and as a result can accurately estimate the size of the existing network, n.

As the new users join the network, the size of the network grows. The network in equilibrium is able to support a certain number of free-riders. Once that level of free-riding is exceeded, congestion becomes too high for all of the users and one or more non-sharer needs to turn their sharing on and a new equilibrium is established. The critical number of sharers k_r in a network of size n satisfies

$$\frac{1}{f\left(\frac{n+1}{k_r+1}\right)-f\left(\frac{n+1}{k_r}\right)} \le \frac{v_j}{c_j} \le \frac{1}{f\left(\frac{n}{k_r+1}\right)-f\left(\frac{n}{k_r}\right)} \tag{2}$$

This condition states that when $k = k_r$, the network can no longer support additional non-sharers (above this level either sharer or non-sharers may join). Additional non-sharers would lead to the collapse of the network. Thus, at this point, only sharers will join the network. Non-sharers will choose not to join since they can anticipate that, if they joined, the network would collapse and they would receive no benefit.[1]

Together these two results suggest that free-riding can both exist initially and persist over time as users respond to network conditions through their sharing decisions. This finding is in contrast to recent suggestions that free-riding may lead to the collapse of P2P networks. The reason for this difference is that, in our model, users anticipate the impact of their decision on the performance of the network through their private utility and respond accordingly. What is interesting is that while this response is in accord with users' private utility, it also enhances the public outcome of the game.

4 CONCLUSIONS AND FUTURE DIRECTIONS

P2P networks have emerged as an important medium for the exchange of information goods. However, little is known about the impact of user behavior on the performance of these networks. Our work represents an initial attempt to analytically model the behavior of these networks. We show that while P2P networks exhibit characteristics of public and club goods, they differ in important ways that impact their performance. We also show that in spite of predictions of the collapse of P2P networks in the face of free-riding (e.g., Adar and Huberman 2000), that free-riding can be sustained by these networks in equilibrium.

In future work, we plan to explore how much the performance of these networks could be improved in the absence of free-riding and managerial incentives and mechanisms to limit free-riding. Specifically, we plan to extend current work by analyzing the feasibility and desirability of non-price mechanisms for reducing free-riding. Non-price mechanisms may be important in the context of P2P networks because of the informal nature of the interaction and the difficulty in determining the identity of users.

---

[1]If we believe that some of the previous period non-sharers choose to be sharers in this period, then even some non-sharers can join the network.

5 ACKNOWLEDGEMENTS

We thank participants at the Carnegie Mellon University Graduate School of Industrial Administration and the University of California at Berkeley's School of Information Management and Systems and the Haas School of Business for valuable comments on this research. Financial support was provided by the National Science Foundation through grant IIS-0118767.

6 REFERENCES

Adar, E., and Huberman, B. A. "Free-Riding on Gnutella," *First Monday* (5:10), 2000.

Asvanund, A., Clay, K., Krishnan, R., and Smith, M. "Preliminary Evidence on Peer-to-Peer Sharing of Music Files," paper presented at the Telecommunications Policy Research Conference, Alexandria, VA, October 27-29, 2001.

Buchanan, J. M. "An Economic Theory of Clubs," *Economica* (32:125), 1965, pp. 1-14.

Golle, P., Leyton-Brown, K., and Mironov, I. "Incentives for Sharing in Peer-to-Peer Networks," Working Paper, Stanford University, Palo Alto, CA, 2001.

Hardin, G. "The Tragedy of the Commons," *Science* (162), 1968, pp. 1243-1248.

Krishnan, R., Smith, M. D., and Telang, R. "The Economics of Peer-To-Peer Networks," Working Paper, Carnegie Mellon University, Pittsburgh, PA, 2002.

Appendix A
Derivation of Expected Length of the Queue on Each Node

Without loss of generality, consider the choice of the n^{th} user conditional on all the other n-1 sharing. Let j denote the total number of users demanding the same content. If the n^{th} user allows sharing, j simply follows a binomial distribution with parameters n – 1 and p. The expected number of hits on any of the other n-1 nodes is:

$$E_s = \sum_{j=1}^{n-1} \binom{n-1}{j} \left(\frac{1}{n-1}\right)^j \left(1-\frac{1}{n-1}\right)^{n-1-j} j \tag{A1}$$

If the n^{th} user doesn't allow sharing, the formula becomes complicated because the n^{th} user has $n-1$ choices while all the other users only have $n-2$ choices. We now consider two exclusive cases in calculating the expected number of hits on any of the other n-1 nodes. First demand on any of the $n-1$ nodes includes the n^{th} user's. Second demand on any of the nodes doesn't include the n^{th} user's. We now get expected number of hits on any of the other n-1 nodes:

$$E_{ns} = \sum_{j=1}^{n-1}\left[\frac{1}{n-1}\times\binom{n-2}{j-1}\times\left(\frac{1}{n-2}\right)^{j-1}\times\left(\frac{n-3}{n-2}\right)^{n-(j+1)} + \frac{n-2}{n-1}\times\binom{n-2}{j}\times\left(\frac{1}{n-2}\right)^{j}\times\left(\frac{n-3}{n-2}\right)^{n-(j+2)}\right]j \tag{A2}$$

Let n be the total number of users and k be the number of sharers. From (A1) and (A2), whether the n^{th} user allows sharing or not, the expected hits on any of the other n-1 nodes (expected length of the queue) is increasing with n and decreasing with k. To capture these relationships, we simplify (1) and (2) to $f(\frac{n}{k})$.

Appendix B
Proof of Proposition 1

Given the belief that there are k-1 users sharing, original users will share if and only if:

$$f\left(\frac{n}{k}\right)v_i - c_i \geq f\left(\frac{n}{k-1}\right)v_i.$$

that is,

$$\frac{v_i}{c_i} \geq \frac{1}{f\left(\frac{n}{k}\right) - f\left(\frac{n}{k-1}\right)} \tag{B1}$$

As long as (B1) is satisfied, some original users are willing to share.

To show stability, we need to show $i \in \{SH\}$, and $j \in \{NS\}$ won't deviate unilaterally. For $i \in \{SH\}$, we need:

$$\frac{v_i}{c_i} \geq \frac{1}{f\left(\frac{n}{k}\right) - f\left(\frac{n}{k-1}\right)} \tag{B2}$$

For $j \in \{NS\}$, we need:

$$\frac{v_j}{c_j} \leq \frac{1}{f\left(\frac{n}{k+1}\right) - f\left(\frac{n}{k}\right)} \tag{B3}$$

As long as (B2) and (B3) are satisfied, the equilibrium is stable given the network size n.